

IL LIBRO

Le istituzioni culturali nel *semantic Web*: riflessioni a margine di *Linked data* *per biblioteche, archivi e musei* di Mauro Guerrini e Tiziana Possemato

di Valeria Lo Castro

Mauro Guerrini; Tiziana Possemato, *Linked data per biblioteche, archivi e musei: perché l'informazione sia del Web e non solo nel Web*. Milano: Editrice Bibliografica, 2015. 257 p. ISBN 978-88-7075-830-6

I dati sono noiose scatole marroni, ed è così che ce li immaginiamo, no? Perché i dati di per sé non sono di immediata applicazione, ma in realtà, i dati determinano tantissime cose nelle nostre vite e ciò accade perché c'è qualcuno che prende quei dati e ne fa qualcosa [...] voglio pensare a un mondo in cui tutti abbiano caricato dati sul Web e dunque praticamente qualunque cosa possiate immaginare sia sul Web e chiamare il tutto linked data [...]¹.

Nel 2009 in un famoso *TED talks*, Tim Berners-Lee, l'inventore del World wide web, lanciò l'idea dei linked data, individuata come la nuova tecnologia portante per la costruzione del *semantic Web*. Si tratta di un insieme di *best practices* per la pubblicazione e il collegamento di dati strutturati sul Web. Queste pratiche sono il fondamento dell'evoluzione dal Web di documenti al Web di dati, uno spazio dati globale che connette dati provenienti da differenti domini.

Il *semantic Web* è un'estensione del Web tradizionale, nel quale all'informazione viene attribuito un significato ben definito (viene aggiunto uno strato semantico), al fine di migliorare la cooperazione tra uomini e macchine.

Esso ha conosciuto due fasi di sviluppo: una prima fase, caratterizzata da un approccio *top down*, altrimenti detto dell'«ontologia fondazionale», è riferibile al periodo 1999-2006 e basata sull'idea della costituzione di un modello di conoscenza

VALERIA LO CASTRO, Centro di ateneo per le biblioteche "Roberto Pettorino", Università degli studi di Napoli Federico II, piazza Bellini 59-60, 80138 Napoli, e-mail valeria.locastro@unina.it.

¹ *Tim Berners-Lee e il Web prossimo venturo*, «TED talks», 4 febbraio 2009, <http://www.ted.com/talks/tim_berniers_lee_on_the_next_web?language=it> (ultima consultazione: 30/01/2016).

universale in grado di esprimere i concetti e le relative relazioni, da cui creare i metadati per etichettare le risorse. Questo approccio ha evidenziato tutti i suoi limiti nella difficoltà di raggiungere uno standard ontologico riconosciuto universalmente e definito a priori poiché il Web è un ambiente dinamico e in costante evoluzione e perché risultava difficile esprimere in un linguaggio formale la semantica naturale e soggetta a diverse convenzioni sociali e individuali. La seconda fase, caratterizzata da un approccio *bottom up* e inaugurata nel 2006, ha visto protagonista ancora una volta Tim Berners-Lee che, insieme a Nigel Shadbolt e Wendy Hall, nell'articolo *The semantic Web revisited*² ridefinisce alcuni aspetti del *semantic Web* proponendo l'idea di sviluppare e migliorare le ontologie in maniera collaborativa, grazie all'intervento di comunità di pratica e alle possibilità offerte dal Web 2.0. I linked data, dunque, ereditano la stessa visione del *semantic Web*, ma rappresentano un modo di strutturare i dati che si manifesta più flessibile e più adatto alle caratteristiche del Web, che è un ambiente dinamico e in continua evoluzione. Il messaggio essenziale è che, qualunque sia il formato in cui le persone vogliono i dati, e indipendentemente dal formato in cui sono forniti, si usa il modello RDF (Resource description framework) come mezzo per collegarli.

Il *semantic Web* nasce da un lato dalla necessità di mettere ordine in un ecosistema informativo caratterizzato sempre più da grandi quantità di dati, di rendere i dati e le informazioni maggiormente visibili sul Web e ricercabili dai motori di ricerca generalisti, dall'altro dovrebbe aumentare le potenzialità del Web tradizionale e superare alcune criticità relative alla ricerca di documenti e di informazioni, favorendo la ricerca esplorativa accanto alla ricerca consapevole. I concetti di interoperabilità, tra sistemi e silos, e il concetto di interazione, tra uomini e macchine, diventano centrali.

Dal punto di vista architettonico il *semantic Web* è stato rappresentato nel 2001 da Tim Berners-Lee come una piramide di sette strati, detta *semantic Web stack* e costituita da nove elementi, che sorregge tre tipologie di informazioni: dati, metadati, classi di dati (nelle quali si esprimono le relazioni fra concetti).

La prima parte del libro di Mauro Guerrini e Tiziana Possemato, *Linked data per archivi, biblioteche, musei* è incentrata sulla definizione di *semantic Web* e sulla tecnologia linked data (capitolo 1), sui mattoncini che contribuiscono alla costruzione del *semantic Web*, a partire dai processi di 'identificazione', attraverso gli *uniform resource identifier*, e 'collegamento', attraverso l'adozione del modello logico RDF, (capitolo 2), fino ad arrivare all'utilizzo di vocabolari, ontologie e schemi di metadati (capitolo 3).

Altri tre capitoli sono dedicati alla ricostruzione del *semantic Web stack*, di cui si diceva prima, (capitolo 4), all'analisi dell'intero ciclo di vita dei linked data e ai meccanismi di pubblicazione al fine di garantirne la qualità, le licenze più appropriate per assicurare la trasparenza e per renderli realmente fruibili e riutilizzabili (capitoli 5-7).

Questo libro di Mauro Guerrini e Tiziana Possemato colma una lacuna e risponde a un'attesa vivamente sentita non solo dagli studiosi di biblioteconomia, archivistica e museografia, ai quali esplicitamente si rivolge, ma anche dai molti ricercatori operanti nei più diversi ambiti disciplinari attenti al modo in cui le reti telematiche hanno trasformato le pratiche della ricerca scientifica, le forme di comunicazione del sapere e le modalità di diffusione dei suoi risultati.

² Tim Berners-Lee; Nigel Shadbolt; Wendy Hall, *The semantic Web revisited*, «Intelligent systems IEEE», 21 (2006), n. 3, p. 96-101.

Così si legge nell'introduzione di Roberto Delle Donne al libro. Appare dunque la volontà di allargare i pubblici di riferimento (operazione, questa, già avviata con *Introduzione a RDA* di Guerrini e Bianchini³), dove analogamente si dichiarava che il libro era rivolto a tutte le persone interessate a trattare e mediare l'informazione, ma anche a coloro che fanno ricerca poiché la rete ha modificato il modo in cui si fa scienza.

Nella prefazione i due autori sottolineano da un lato la propensione umana a creare relazioni e formare comunità, dall'altro la necessità di trovare un linguaggio per comunicare. I linked data vengono definiti come il nuovo linguaggio della comunicazione globale del Ventunesimo secolo.

Se nella presentazione al *TED*, Tim Berners-Lee sottolinea la potenza dei dati soprattutto in relazione al loro collegamento (*data are relationship*), in esergo a questo libro i due autori hanno scelto di riportare un altro passo di Tim Berners-Lee, questa volta riferito ad una questione cruciale e particolarmente alla necessità di avere una struttura per questi dati:

In una visione estrema, il mondo può essere visto soltanto come connessioni, nient'altro. Noi pensiamo a un dizionario come a un archivio di significato, ma esso definisce parole soltanto con altre parole. Mi piaceva l'idea che un pezzo di informazione fosse davvero definito solo da cosa è collegato a, e come. Non c'è nient'altro da capire. La struttura è tutto.

Il libro, per ritornare alle parole di Delle Donne, incentrato sui principi fondamentali dei linked data rappresenta un importante contributo di analisi e di riflessione storica, teorica, tecnica e funzionale al fine di delineare «modalità e strategie di integrazione di fonti informative eterogenee in un'unica piattaforma di ricerca». Alla base di questi studi, il modello e la filosofia che ne sorregge l'ispirazione è senz'altro quella di voler promuovere la cultura dell'apertura dei saperi e della scienza.

Guerrini e Possemato declinano questo paradigma e questa visione agli oggetti e alle istituzioni culturali. Il lavoro è il risultato di un lungo cammino, iniziato con la pubblicazione di un primo contributo che ha inaugurato in Italia la riflessione sull'argomento, proprio a firma di Guerrini e Possemato su *Biblioteche oggi*, *Linked data: un nuovo alfabeto del web semantico*⁴ e da un importante convegno internazionale tenutosi a Firenze "Global interoperability and linked data in libraries"⁵, dove sono stati presentati, tra l'altro, i casi più interessanti di applicazione dei linked data in grandi biblioteche nazionali. Articoli e contributi di diverso tipo, nel corso del tempo, sono gradualmente aumentati e hanno arricchito la riflessione biblioteconomica, allargando sempre più la platea degli studiosi e dei lettori, seguiti dalle prime sperimentazioni pratiche (linked open data in SBN, o il consorzio costituito da una rete di atenei italiani, ossia Università di Napoli Federico II, Napoli L'Orientale, Napoli Parthenope, Salerno, Sannio, Basilicata, Salento, che sta convertendo e integrando i cataloghi delle diverse istituzioni in linked open data).

La seconda parte del lavoro è più incentrata sulle strutture di applicazione dei linked data e dunque su biblioteche, archivi e musei: sulla riflessione teorica relativa

3 Carlo Bianchini; Mauro Guerrini, *Introduzione a RDA*. Milano: Bibliografica, 2014.

4 Mauro Guerrini; Tiziana Possemato, *Linked data: un nuovo alfabeto del Web semantico*, «Biblioteche oggi», 30 (2012), n. 3, p. 7-15.

5 V. *Global interoperability and linked data in libraries: special issue*, «JLIS», 4 (2013), n. 1, <<http://leo.cineca.it/index.php/jlis/issue/view/536>>.

ai formati da adottare (capitolo 8) e in particolare sulla necessità soprattutto per i cataloghi di essere riconvertiti e di utilizzare il modello logico RDF e altri schemi di metadati come RDA (Resource description and access), individuato dai due autori come lo schema ideale per l'evoluzione e il transito verso i linked data, su BIBFRAME, modello di dati ispirato a FRBR elaborato dalla Library of Congress, per lo scambio dei dati bibliografici nel *semantic Web* in contesti diversi; sui *case studies* relativi al lavoro delle Biblioteche nazionali e delle *digital libraries* (capitolo 9).

Assai interessante, infine, il capitolo finale relativo ai vantaggi per l'utenza che utilizzerà i dati.

La conversione dei cataloghi in linked open data attualmente si manifesta attraverso due strade: quello della messa disposizione di *dataset* grezzi per gli sviluppatori per un uso non previsto né prevedibile a priori e la costruzione di un portale 'enciclopedico', sempre meno legato alla singola opera di un autore e sempre più risultato di un lavoro a tutto tondo sul personaggio e il complesso delle sue opere.

Il primo effetto del trasferimento dei dati dai silos bibliografici al Web dovrebbe essere che la registrazione di un oggetto d'interrogazione sarà raggiungibile indifferentemente a partire da una *query* sul Web con un motore di ricerca o su un OPAC di un servizio bibliotecario: in entrambi i casi l'utente dovrebbe essere in grado di compiere ricerche consapevoli e ricerche esplorative. La differenza importante tra le due modalità consisterà nel fatto che i linked data creati dalle biblioteche saranno costituiti e presentati secondo standard di qualità e largamente condivisi, offrendo una maggiore garanzia di coerenza e autorevolezza. Il secondo effetto dovrebbe essere la ricerca di un dato a partire da qualsiasi forma sia nota; un utente di qualsiasi parte del mondo e di qualsiasi lingua potrebbe eseguire interrogazioni usando termini nella forma da lui preferita o conosciuta e trovare tutte le informazioni rilevanti sulle manifestazioni.

Concludono il volume un saggio di Carlo Bianchini sull'evoluzione dei cataloghi, *Dai cataloghi alla navigazione semantica*, un'appendice e un glossario dei termini tecnici. Da sottolineare anche l'apporto di altri studiosi nella redazione di alcuni parti del testo, dell'appendice, nella lettura, a testimonianza di un lavoro in cooperazione.

Questo libro invita a entrare nell'ottica di un uso di linguaggi e standard comuni alle diverse comunità disciplinari presenti sul Web e invita i bibliotecari a staccarsi dalla prospettiva della propria comunità di riferimento per interagire e a cooperare con le altre, quindi 'cooperazione' è la prima parola chiave, 'apertura' è la seconda, poiché linked data non è altro che il risvolto tecnologico che sorregge la volontà di promuovere la cultura dell'apertura dei saperi e della scienza. Altro aspetto è quello di far emergere il potenziale delle macchine a fronte della mole di informazioni da processare e trattare. Internet è un luogo dove interagiscono umani e macchine, è necessario sfruttare di più e meglio la capacità e la potenza di calcolo di queste ultime. La comunità del *semantic Web* ha avvertito l'importanza e il valore dei dati prodotti dalle istituzioni del patrimonio culturale, informazione ricca, strutturata, controllata e autorevole, ma contenuta in silos spesso non interoperabili, ma il rapporto è bilaterale poiché anche gli istituti della memoria registrata sono interessati alle tecnologie del *semantic Web*, al fine di integrare fonti informative eterogenee in piattaforme che consentano di interrogare i dati del patrimonio culturale, ma anche prendere parte attiva allo sviluppo del Web nell'ottica dell'integrazione dei saperi e delle conoscenze scientifiche.

Le istituzioni culturali hanno da sempre raccolto dati relativi alla loro attività. Si tratta di informazione spesso altamente strutturata, contenuta in diversi ambienti o silos. In molti casi le biblioteche e i bibliotecari hanno contribuito alla creazione di questi depositi istituzionali, mettendo a frutto le competenze acquisite in relazione

allo sviluppo delle collezioni, alla gestione dei metadati e dei servizi di indicizzazione e recupero dell'informazione. Negli ultimi anni le biblioteche hanno progressivamente riconosciuto gli strumenti del *semantic Web* come adeguati ai propri contenuti, e ne hanno compreso le potenzialità nel favorire una diffusione molto più ampia di dati e metadati bibliografici standardizzati, nonché il loro riuso in contesti anche diversi da quelli di origine e nel garantire una loro maggiore visibilità sul Web. Alcune considerazioni finali: intanto la necessità di costruire, nel momento in cui si sceglie di dar vita a progetti di questo tipo, gruppi di lavoro con diverse competenze. L'apporto dei bibliotecari risulta prezioso nella misura in cui essi possono mettere a disposizione le loro competenze per garantire la qualità e l'autorevolezza dei *datasets* e dei vocabolari. Anche il lavoro di descrizione che fin qui ha notevolmente impegnato i bibliotecari potrebbe diminuire in favore dei processi di identificazione e collegamento. Portare i propri dati strutturati in questo mondo significa valorizzarli e dunque in qualche modo mettere a frutto il lavoro che le biblioteche hanno fatto in tutti questi decenni passati, nonché creare le condizioni effettive per quella cooperazione di cui nel settore bibliotecario si parla da tempo, ma che finora si è sviluppata in modo per certi versi limitato.

Articolo proposto il 19 dicembre 2015 e accettato il 21 dicembre 2015.

ABSTRACT

AIB studi, vol. 56 n. 1 (gennaio/aprile 2016), p. 111-115. DOI 10.2426/aibstudi-11437

VALERIA LO CASTRO, Centro di ateneo per le biblioteche "Roberto Pettorino", Università degli studi di Napoli Federico II, piazza Bellini 59-60, 80138 Napoli, e-mail valeria.locastro@unina.it.

Le istituzioni culturali nel *semantic Web*: riflessioni a margine di *Linked data per biblioteche, archivi e musei* di Mauro Guerrini e Tiziana Possemato

Il contributo discute – a partire dal libro *Linked data per biblioteche, archivi e musei* – i benefici che le istituzioni culturali come, appunto, biblioteche, archivi e musei possono trarre dall'utilizzo degli strumenti del *semantic Web*, e riflette sull'apporto che esse possono dare all'intera comunità del Web. Il libro di Mauro Guerrini e Tiziana Possemato presenta in maniera organica e sistematizzata i principi di base relativi ai linked data, introduce gli standard, le tecnologie e i linguaggi per la loro creazione e gestione, ne descrive il ciclo di vita e presenta le iniziative e i progetti nazionali e internazionali più rilevanti, in modo da consentire al lettore di comprendere e interpretare i linked data e le problematiche legate alla loro produzione, pubblicazione e utilizzo.

Cultural institutions in the semantic Web: considerations about *Linked data per biblioteche, archivi e musei* by Mauro Guerrini and Tiziana Possemato

This paper – starting from the book *Linked data for libraries, archives and museums* (Mauro Guerrini and Tiziana Possemato, 2015) – discusses how cultural institutions, such as libraries, archives and museums, could benefit from the semantic Web, and how they could contribute to the Web community. The book presents linked data basic principles, suggesting standards, technologies and languages for their creation and management. Describing their life cycle, as well as the main national and international projects, it enables the reader to understand what linked data are, and to figure out some critical issues related to their production, publication and use.